

**Claims**

1. A method of processing a continuous audio stream containing human speech related to at least one particular transaction, comprising the steps of:
  - digitizing the continuous audio stream;
  - 5 detecting a speaker change in the digitized audio stream;
  - performing a speaker recognition if a speaker change is detected;
  - transcribing at least part of the continuous audio stream if a predetermined speaker is recognized.
2. A method of processing a continuous audio stream containing human speech related to at least one particular transaction, comprising the steps of:
  - digitizing the continuous audio stream;
  - detecting a speaker change in the digitized audio stream;
  - 10 performing a speaker recognition if a speaker change is detected;
  - indexing the audio stream with respect to the detected speaker change if a

predetermined speaker is recognized.

3. Method according to claim 1 or 2, comprising the further step of protocolling time information for detected speaker changes.

4. Method according to any of the preceding claims, wherein the step of detecting  
5 a speaker change and/or the step of performing a speaker recognition is/are preceded by  
the further step of detecting non-speech boundaries between continuous speech segments.

5. Method according to any of the preceding claims, wherein the step of detecting  
a speaker change is accomplished by use of at least one characteristic audio feature, in  
particular features derived from the spectrum of the audio signal.

10 6. Method according to claim 1 or 2, wherein the step of performing a speaker  
recognition involves the particular steps of calculating a speaker signature from the audio  
stream and comparing the calculated speaker signature with at least one known speaker  
signature.

7. Method according to any of the preceding claims for use in a speech  
15 recognition or voice control system comprising at least two speaker-specific speaker  
models and/or dictionaries, wherein interchanging between the at least two  
speaker-specific dictionaries dependent on the detected speaker change and the

corresponding recognized speaker.

8. Apparatus for processing a continuous audio stream containing human speech related to at least one particular transaction, comprising:

means for predetermining at least one speaker;

5 means for detecting speaker changes in the audio stream;

means for recognizing the predetermined speaker in the audio stream;

means for initiating transcription of at least part of the audio stream in case of a detected speaker change and a recognized predetermined speaker.

9. Apparatus for processing a continuous audio stream containing human speech  
10 related to at least one particular transaction, comprising:

means for predetermining at least one speaker;

means for detecting speaker changes in the audio stream;

means for recognizing the predetermined speaker in the audio stream;

means for indexing the audio stream dependent on a detected speaker change and  
15 a recognized predetermined speaker.

10. Apparatus according to claim 8 or 9, further comprising means for detecting non-speech boundaries between continuous speech segments.
11. Apparatus according to any of claims 8 to 10, further comprising means for automatically scanning a continuous audio record, in particular a continuous audio stream  
5 recorded on a data or a signal carrier, and for detecting speaker changes in the continuous audio record.
12. Apparatus according to any of claims 8 to 11, further comprising means for continuously monitoring a real-time continuous audio stream and performing the steps of claim 1 or 2.
- 10 13. Apparatus according to any of claims 8 to 12, further comprising log means for protocolling time information for the at least one detected speaker change.
14. Apparatus according to any of claims 8 to 13, comprising means for marking at least the beginning of a detected speech segment related to a predetermined speaker.
15. Apparatus according to any of claims 8 to 14, comprising data base means for  
15 storing speech signatures for at least two speakers.
16. Speech recognition or voice control system processing an incoming audio

- stream and having at least two speaker models and/or speaker-specific dictionaries, comprising means for detecting a speaker change in the incoming audio stream; means for gathering speaker-specific information and for comparing the gathered speaker-specific information with corresponding speaker-specific information of at least one predetermined speaker thus recognizing the at least one predetermined speaker; means for interchanging between the at least two speaker-specific dictionaries dependent on the detected speaker change and the corresponding recognized speaker.
17. A data processing program for execution in a data processing system comprising software code portions for performing a method according to any of claims 1 to 7 when said program is run on said computer.
18. A computer program product stored on a computer usable medium, comprising computer readable program means for causing a computer to perform a method according to any claims 1 to 7 when said program is run on said computer.